



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Knowledge and Cognitive Integration

**Citation for published version:**

Palermos, S 2014, 'Knowledge and Cognitive Integration', *Synthese*, vol. 191, no. 8, pp. 1931-1951.  
<https://doi.org/10.1007/s11229-013-0383-0>

**Digital Object Identifier (DOI):**

[10.1007/s11229-013-0383-0](https://doi.org/10.1007/s11229-013-0383-0)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Synthese

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



(forthcoming in *Synthese*)

# KNOWLEDGE AND COGNITIVE INTEGRATION

S. Orestis Palermos

University of Edinburgh

**Abstract:** Cognitive integration is a defining yet overlooked feature of our intellect that may nevertheless have substantial effects on the process of knowledge-acquisition. To bring those effects to the fore, I explore the topic of cognitive integration both from the perspective of virtue reliabilism within externalist epistemology and the perspective of extended cognition within externalist philosophy of mind and cognitive science. On the basis of this interdisciplinary focus, I argue that cognitive integration can provide a minimalist yet adequate epistemic norm of subjective justification: so long as the agent's belief-forming process has been integrated in his cognitive character, the agent can be justified in holding the resulting beliefs merely by lacking any doubts there was something wrong in the way he arrived at them. Moreover, since both externalist philosophy of mind and externalist epistemology treat the process of cognitive integration in the same way, we can claim that epistemic cognitive characters may extend beyond our organismic cognitive capacities to the artifacts we employ or even to other agents we interact with. This move is not only necessary for accounting for advanced cases of knowledge that is the product of the operation of epistemic artifacts or the interactive activity of research teams, but it can further lead to interesting ramifications both for social epistemology and philosophy of science.

## 1. INTRODUCTION

Cognitive integration is an overlooked yet defining feature of our intellect. It is the reason why, in contrast to mere stimulus-response automata, we entertain advanced beliefs in an epistemically responsible way, and why we can do so even in the complete absence of any reasons to back those beliefs up.

Nevertheless, within philosophy of mind and cognitive science, it is only recently that the topic of cognitive integration was brought into focus. This recent change of focus, however, has not been and is still not guided by an attempt to understand how our complex brain capacities intertwine with each other. Due to Fodor's (1983) persuasive understanding of our intellectual

architecture as modular, in-the-head cognition has largely thought to be for the most part domain specific and informationally encapsulated.<sup>1</sup>

Consequently, cognitive scientists who take the brain to be the primary cognitive explanandum have so far been indisposed to explore and understand the phenomenon of cognitive integration.

Instead, only after the hypothesis of extended cognition (Clark & Chalmers 1998) was proposed (according to which external elements can be proper parts of our cognitive systems), did the topic of cognitive integration start to attract attention; claiming that artifacts can be parts of an agent's cognitive system presupposes an account of how such external elements can be properly integrated into our cognitive loops. Accordingly, over the past fifteen years, several attempts have been made to account for the process of cognitive integration, most of which rely either on a common-sense functionalist understanding of our minds (Clark & Chalmers 1998; Clark 2010), or on a mathematically inspired approach (Clark 2008; Chemero 2009, Palermos 2014) that focuses on the dynamical nature of cognitive processes.

Interestingly, however, the topic of cognitive integration has also recently emerged within epistemology in reference to the concept of our epistemic *cognitive characters*. Even within epistemology, however, the discussion of cognitive integration has not been entirely unrelated to the idea of cognitive extension. Admittedly, Greco (2010)—the first to discuss (in passing) the idea of cognitive integration in epistemological terms—does not commit himself to the possibility of cognitive extension. Both Pritchard (2010) and I (Palermos 2011), however, have noted that the notion of the epistemic agent's *cognitive character*, to which all of the agent's knowledge-conducive belief-forming processes must have been properly integrated, is open to an interpretation along the lines suggested by the extended cognition hypothesis. However, and despite these few efforts to elucidate the process of cognitive integration within epistemology, more needs to be said since the effects of this process on knowledge could turn out to be surprisingly substantial.

To bring these effects to the fore, we need to approach the idea of cognitive integration both from the perspective of virtue reliabilism in externalist epistemology (section 2.2) and the perspective of extended

---

<sup>1</sup> Despite the influence of Fodor's work within philosophy of mind and cognitive science, his modular understanding of the mind has met some considerable resistance by equally influential philosophers. One clear example is the work of Churchland (1979, 1988, 1989).

cognition in externalist philosophy of mind (section 2.3). This interdisciplinary focus will help us make apparent, in section 2.3, that both views treat cognitive integration in essentially the same way—an observation that we can then use to bring both perspectives together and thereby demonstrate how our epistemic cognitive characters can extend beyond our organismic cognitive capacities to the epistemic artifacts we employ. This combination of externalist epistemology with externalist philosophy of mind, however, as I will argue in section 3, is not just an available option, but is actually necessary for accounting for advanced cases of knowledge whereby one's true believing is the product of the operation of epistemic artifacts. Furthermore, this approach can generate interesting ramifications both for social epistemology and philosophy of science, wherein the pursuit of knowledge has been traditionally associated with the use of artifacts, carefully tailored labs, and the combined efforts of epistemic agents working in research teams. To start with, however, a few introductory remarks are first in order.

## 2. THE PROCESS OF COGNITIVE INTEGRATION

### *2.1 Introduction*

"John is sad" is an assertion that can be true or false. In producing this claim, of course, perception will always be foundational in a certain way, but in order to make, as well as check, the validity of the claim we combine several other processes as well. In the background of our minds we have a biological as well as a primordial (animalistic) conception of what a human being is and we also have a socio-contextual sense of who "John" is; we have a theory of mind that enables us to understand other people with thoughts and emotions like ours, and we may even need a more personal theory of John's character: John has spent the entire night socializing at the pub and he just made a witty joke, but the downwards motion of his eyes accompanied by a melancholic smile at the end of his remark are enough to indicate to his close friends that his mind is secretly occupied with his recent loss.

It is in a sense like this that theories (commonsensical as well as scientific ones) have a top-down effect on the bottom-up input we receive

from our sense-apparatuses. This overriding effect, known as the theory-ladenness of observations, has been well-studied both within philosophy of science (Kuhn 1962; Hanson 1961; 1969) and philosophy of mind (Churchland 1979; 1988; 1989; Fodor 1984; 1988), and indicates that perceptual beliefs are never ‘purely perceptual’, but they instead emerge out of the *cooperation* of several intellectual capacities operating in tandem.

Now, considerations like these indicate that perception (traditionally thought of as the most foundational aspect of our belief systems) may not be all that basic, and have thereby hinted towards a relativistic picture of science—and possibly epistemology as well—whose most prominent proponents are thought to be Feyerabend (1975) and Kuhn (1962) (the latter quite possibly unjustly though). Fortunately, however, at least as far as the theory-ladenness of observation is concerned, recent studies within cognitive psychology not only point away from relativism, but they also demonstrate how the *interaction* between our theoretical beliefs and our sensory apparatus has a facilitatory effect on the overall process of perception. For example, exploring the literature on relevant experiments, Brewer and Lambert (2001) concede that “perception is determined by the *interaction* of top-down theory information and bottom-up sensory information” (178, my emphasis):

However, note that in all of the above cases the stimuli were either ambiguous, degraded, or required a difficult perceptual judgment. In these cases the weak bottom-up information allowed the top-down influences to have a strong impact on perceptual experience. It seems likely that strong bottom-up information will override top-down information. [...] Thus, the topdown/ bottom-up analysis allows one to have cases of theory-laden perception, but does not necessarily lead down the slippery slope of relativism (*ibid.*).<sup>2</sup>

---

<sup>2</sup> Similarly, Estany (2001, 208) holds that

the beliefs of the higher or more fundamental level influence how perceptual units are interpreted by the lower levels [...] Humans use both types of processes in perception because each have characteristic advantages and disadvantages. Thanks to top-down processes we can recognize patterns with incomplete or degraded information. Moreover, top-down processes make perception faster, but they can induce us to make mistakes in a perception by relying on previous knowledge.

Nevertheless, Estany further notes that even though our perceptual systems get guidance from higher-order expectations, when attention is caused by the mismatches between expectation and reality, the inputs from the arousal system constitute a “reset wave” making it possible to avoid arbitrary relativistic errors of perception (Estany 2001, 213).

Therefore, we should neither be misguided nor feel intimidated by the possibly false impression of relativistic bias that the integration of our intellectual capacities into one overarching belief-generating system may have initially created. To the contrary, we should focus on the phenomenon of cognitive integration because, as we shall see, it actually has a very important facilitatory effect on our epistemic standing. In particular, the interconnectedness of our cognitive capacities allows us to be subjectively justified even if we lack explicit reasons for holding our beliefs. Provided that we act in a conscientious mode—i.e., provided we are motivated to believe what is true—cognitive integration allows us to trust the deliverances of our cognitive abilities by merely lacking any negative reasons against (as opposed to possessing positive reasons for) our beliefs. This sense of epistemically adequate—yet unreflective—cognitive responsibility can only be achieved by agents like us, whose intellectual capacities are appropriately interconnected such that in cases where there is something wrong with the way we form our beliefs or with the beliefs themselves, we will be able to notice this and respond appropriately. Otherwise—if there is nothing wrong—we can go on about with our daily activities without questioning our epistemic standing with respect to every single of the millions (possibly billions?) of beliefs we enjoy in the course of our days.

On a first pass, this probably sounds sketchy, but focusing on contemporary epistemology should allow us to both understand what this unreflective sense of cognitive responsibility amounts to and why it is so important.

## 2.2 Cognitive integration in epistemology

To start with, consider epistemic internalism, which takes an approach to epistemic responsibility that is very different from the one suggested here. According to traditional forms of internalism, one should always be able, at least in principle, to access the reasons that justify one's beliefs, by *reflection alone*.<sup>3</sup> This may initially sound as a reasonable demand, but the problem is that it creates serious complications with respect to our perceptual and empirical beliefs. Specifically, it poses the requirement that there be necessary

---

<sup>3</sup> For classical defenses of this view see Chisholm (1977) and Bonjour (1985, ch. 2). See also Steup (1999), Pryor (2001, §3), Bonjour (2002), Pappas (2005), and Poston (2008).

support relations between one's empirical and perceptual beliefs and one's evidence for holding them (such that one can be in a position to justify one's empirical and perceptual beliefs by reflection alone). As Hume's problem of induction demonstrates, however, this is impossible. Accordingly, it has been traditionally assumed that Hume's arguments lead to skepticism about our empirical knowledge.<sup>4</sup>

Contemplating on the Humean problematic, however, Greco (1999) argues that this is too fast. Hume's arguments should not be directly considered as skeptical ones. Instead, the immediate conclusion to be drawn from them is that there are no necessary support relations between our empirical beliefs and their evidence; that if the evidence for our empirical beliefs is reliable, then it is at most contingently reliable. This realization alone, however, cannot automatically lead to skepticism. Only after we embrace the internalist understanding of knowledge, such that there always be necessary support relations between one's evidence and one's beliefs do we face skepticism.

In other words, in order to avoid skepticism about empirical and perceptual knowledge, we must allow knowledge to be grounded on evidence that is merely contingently reliable, and so we must give up the requirement that one's beliefs should always be internally—i.e., by reflection alone—justified. Any adequate epistemology must be able to account for the fact that *merely contingently reliable evidence can give rise to knowledge* (Greco 1999, 273).

Now, in order to accommodate the above realization, contemporary epistemologists have put forward process reliabilism; *viz.*, the idea that knowledge is true belief that is the product of reliable belief-forming processes, where a reliable process is a process that results in a

---

<sup>4</sup> The problem of induction is well known. We form our beliefs about unobserved matters of fact and the external world on the basis of evidence provided by past and present observations and sensory appearances, respectively. In order, however, for the support relations between our empirical and perceptual beliefs and the evidence offered in their support to be necessary, we also need the further assumptions that the future will resemble the past and that sensory appearances are reliable indications to reality, respectively. The problem, however, is that both of these assumptions rely for their support on what they assert. Consequently, given that circular reasoning is invalid, there are no necessary support relations between our empirical beliefs and the evidence offered in their support. Accordingly, the conclusion that has been traditionally drawn is that our empirical and perceptual beliefs cannot amount to knowledge. For more details on a reconstruction of Hume's skepticism along these lines, see (Greco 1999).

preponderance of true over false beliefs. Moreover, in direct response to the Humean problematic, this approach “denies that one must know that one’s evidence is reliable”, by making “*de facto* reliability the grounds of positive epistemic status” (Greco 1999, 284-5). Accordingly, process reliabilism is an externalist approach to knowledge, because—contrary to the traditional account of knowledge as internally justified true belief—on this view, in order to know, one does not need to know, be justified in believing (by reflection alone, or any other means), or even believe that one’s beliefs are formed in a reliable fashion. So long as one employs an objectively reliable process, one is justified in holding the resulting belief.

Process reliabilism, therefore, has the resources to overcome the Humean skepticism. There are, however, two serious complications with the view. The first one is that process reliabilism, as it stands, is too weak a condition on knowledge because it allows *any* reliable belief-forming process to count as knowledge-conducive, and this, as we shall see, is intuitively incorrect. The second complication is that by making “*de facto* reliability the grounds of positive epistemic status” process reliabilism misses a very important dimension of our epistemic nature. While it is true that in order to know we do need the way of forming our beliefs to be objectively reliable, this sort of objective justification is not sufficient in its own. What we further need is that we be subjectively justified in the sense that we must be somehow sensitive to the reliability of our evidence.<sup>5</sup> Process reliabilism, however, ignores this dimension of our epistemically sentient nature altogether, to the extent that it has been even criticized that it equates us to mere stimulus-response automata (Fuller 2012).<sup>6</sup> We can better appreciate these two problems by taking a look at a few (eccentric, yet informative) examples. Consider Hercules first:

---

<sup>5</sup> Remember, however, that if, as Hume’s skeptical arguments demonstrate, the relation between evidence and belief is not necessary (see also fn. 4), then it is far from obvious how a person can be subjectively justified, especially in externalist approaches such as process reliabilism. If a condition of ‘subjective sensitivity to the reliability of one’s evidence’ must be satisfied, then this should better be accomplished in a way that will not require *knowledge of or even beliefs about* the said reliability (otherwise Hume’s skepticism will strike back).

<sup>6</sup> “Epistemic zombies” would probably be the name that David Chalmers would give to such creatures. Given, however, the present discussion I don’t think they could really exist.



Hercules (Adapted from Pritchard's Temp (2009, 48))

Hercules tosses a drachma whenever he wants to form a belief about the weather outside. If it is heads, he forms the belief that it is sunny; if it is tails he believes it is cloudy; and if it balances in between, he believes it is rainy. As it happens, Hercules' way of forming his weather beliefs is perfectly reliable, because Zeus, who wants to save Hercules from the embarrassment of forming false weather beliefs, has an eye on him; every time he sees Hercules tossing the coin arranges the world accordingly.

Hercules' beliefs are formed in a highly reliable way. So, according to process reliabilism, Hercules has knowledge of the weather conditions. Intuitively, however, this is incorrect. There is a problem with the direction of fit between his beliefs and the facts. In cases of knowledge, we want our beliefs to be true because they correspond to the facts, and not because the facts comply with our beliefs; when one knows, one's true beliefs are about the world, not the other way around. In Hercules' case, however, his beliefs are not true because they are formed in a way that detects the facts. Instead, he first forms his beliefs in an arbitrary way—he makes no efforts to ensure they will come out true—and then Zeus takes over so that the facts will comply with Hercules' beliefs. This, however, is not knowledge; it is the 'luck of the gods'. If one day Zeus had a fight with Hera, Hercules' beliefs would cease coming out true.

Notice, however, that if Hercules used his cognitive abilities—say by taking a look at the sky—to form his weather beliefs, then he would not run into any such problems. If he didn't form his beliefs in an arbitrary way, but on the basis of his cognitive abilities, he would not need Zeus to tweak the world so that his beliefs could systematically turn out true. If one's beliefs are the product of one's cognitive abilities, then if they turn out to be true it will be because they are sensitive to the facts; the direction of fit will be the correct one.

So, it may be proposed that the way to restrict the reliable belief-forming processes to those that get the direction of fit correctly—such that they can be knowledge-conducive—is to identify them with one's cognitive abilities, or, in other words, with those processes that can be intuitively thought of as cognitive ones. But can all *prima facie* cognitive processes count as cognitive abilities and thereby produce knowledge? The answer, as we shall now see, must be a negative one, because there are reliable processes that we might be inclined to categorize as cognitive ones, but which fail to

deliver knowledge, exactly because they disallow the agent to be *subjectively* justified in employing them. Think about the Serendipitous Brain Lesion first:

Serendipitous Brain Lesion (Greco 2010, 149)

Suppose that *S* has a rare brain lesion, one effect of which is to reliably cause the true belief that one has a brain lesion. Even if the process is perfectly reliable, it seems wrong that one can come to have knowledge that one has a brain lesion on this basis.

Again, the unfortunate agent's way of forming his true belief about the brain lesion on the basis of his brain lesion is reliable. Process reliabilists, therefore, must accept that he can gain knowledge in this way. As Greco claims, however, this does not sound correct. Why not? Mainly because the way the agent forms his belief is so strange from his point of view that he cannot accept he can gain knowledge in this way (Greco 1999, 2010). Accordingly, there might be reliable in-the-head processes (such that we may be inclined to call them cognitive ones) that we wouldn't like to claim they are knowledge-conducive cognitive abilities, because from the agent's point of view they are strange. More precisely, the underlying intuition here is that for a process to be eligible to count as a cognitive ability it must not be strange, in the sense that it must not be at odds with the rest of the agent's cognitive system.<sup>7</sup> The reason is that if the process is strange, then, in light of the rest of his cognitive system, the agent will reject both the process and its deliverances despite the fact that they are in fact reliable—from the agent's point of view, they aren't. So, in order for a process to be a candidate for qualifying as a cognitive ability such that it can be knowledge-conducive it must not be inconsistent with the rest of the agent's beliefs and his methods of producing them. In other words, it must be such that it can become part of, or be integrated into, the rest of the

---

<sup>7</sup> An anonymous referee points out that strangeness is description-relative. Take vision for example. We are all familiar with acquiring knowledge through seeing things. But learning about the physiological and neural underpinnings of vision will surely seem strange to some; couldn't such a person say "This is really strange, and I don't really see how it works, but, I guess, this is how I know the color of my shoes"? I think this is right, but this example wouldn't be problematic for the following two reasons. First, even though the explanation may seem 'strange' to the agent (in the sense of being difficult to understand) it is not 'at odds with the rest of the agent's cognitive system'. Second, the requirement that the process not be strange does not refer to a reflective-explanatory understanding of the process (as in the referee's example), but to the presence of the process itself. Think about the analogy of a strange (i.e., eccentric) person who is nevertheless not a stranger: the requirement that the process not be strange allows for the process to be 'strange' in the first sense, but not in the sense of being a 'stranger'. I am thankful to the referee for bringing this ambiguity to my attention.

agent's cognitive system (Greco 2010, 152). And clearly, this is not the case with the serendipitous brain lesion. The process is a cognitive malfunction, and even more crucially, its output is so odd that no epistemic agent could accept as true. In other words, the serendipitous brain lesion cannot count as knowledge-conducive because it is so strange that it cannot become part of the rest of the agent's cognitive system and so cannot count as a cognitive ability.<sup>8</sup>

Nevertheless, even a reliable process that is normal enough to become part of the agent's cognitive system cannot yet count as a cognitive ability that can produce knowledge. Consider a further example:

Careless Math Student (Greco 2010, 149)

Suppose that *S* is taking a math test and adopts a correct algorithm for solving a problem. But suppose that *S* has no understanding that the algorithm is the correct one to use for this problem. Rather, *S* chooses it on a whim, but could just as well have chosen one that is incorrect. By hypothesis, the algorithm is the right one, and so using it to solve the problem constitutes a reliable process. It seems wrong to say that *S* thereby knows the answer to the problem, however.

The careless student's algorithm for solving the problem is also reliable. But again, we cannot attribute knowledge to her. Why not? The reason is that she employed the right method on a whim, such that she could have very easily employed another, incorrect method. Her reliable process is a fleeting one. It is not a habit or a disposition of hers. Given the same circumstances, she could have employed an inappropriate method, thereby, ending up with a falsehood. If, however, the student had habitually invoked the correct

---

<sup>8</sup> An anonymous referee insists that the brain lesion can yield knowledge, thereby implying that process reliabilism is a sufficient condition on knowledge. While it may be possible to make the case for the sufficiency of process reliabilism, the orthodox view within mainstream epistemology goes against this prospect. For classical rejections of the sufficiency of process reliabilism on the basis of thought experiments very similar to the Serendipitous Brain Lesion, see Bonjour (1980), Lehrer (1990) and Plantinga (1993b). The main idea is that reliability might be necessary for knowledge but what is further required is satisfaction of the internalist intuitions with respect to the possession of subjective justification (as we mentioned above, one of the problems for process reliabilism is that by making *de facto* reliability the grounds of positive epistemic status, it fails to capture the intuition that, somehow, we must also be sensitive to the reliability of our evidence). Internalists, typically require the possession of reflectively accessible reasons for said reliability. Here, following Greco's intuitions (1999; 2010, 149-155), we opt for a weaker condition of subjective justification according to which the agent must lack beliefs against the reliability of his belief-forming process, where the process being strange from the agent's point of view would count as just one such defeating belief. For very similar intuitions on how the strangeness of the origin of the relevant beliefs acts as a defeater in Bonjour (1980) and Lehrer's (1990) thought experiments see Goldman (1986, 111-112).

algorithm when the problems called for it, then we would indeed be inclined to claim that she could gain knowledge on its basis. The reason for this is that if a process is a disposition or a habit of the agent, then the agent will be able to become aware of the circumstances in which it can become unreliable. Otherwise, it seems arbitrary that the agent employed it in an appropriate, but isolated case, and so cannot gain knowledge on its basis. In other words, a reliable process that is normal—such that it can, in principle, become part of the rest of one’s cognitive system—won’t be a candidate for qualifying as a cognitive ability, unless it is also a disposition or a habit of the agent. Why is this so? The intuition is that abilities, in general, are habits or dispositions possessed by agents.<sup>9</sup> But apart from such intuitions we have also noted that in order for a reliable process to count as a cognitive ability it must be such that it can become part of (or be integrated into) the rest of the agent’s cognitive system. One requirement for this, we have noted, is that the process not be strange such that it won’t be inconsistent with the rest of the agent’s cognitive system. What is further required, however, is that it also be coherent with her cognitive system in the following sense: The agent must be able to become aware that the process is *unreliable* in certain circumstances, because this will allow her to *non-accidentally* endorse its deliverances in the rest of the circumstances, *even if she lacks any positive beliefs for its reliability*. And in the absence of any explicit reasons that are accessible through reflection alone—recall the Humean problematic—the only realistic way for the agent to so

---

<sup>9</sup> An anonymous referee is worried that I should not use ‘dispositions’ and ‘habits’ as synonyms. Specifically, not all dispositions are habits; someone or something, for example, may be disposed to act in a certain way—should the appropriate conditions obtain—even if the relevant person or thing has never behaved in that way before. Accordingly, the worry further goes, the fact that a cognitive ability is a disposition does not mean it will also be a habit. In response, even though it is true that in one sense of the term, ‘dispositions’ are not always going to be habits, there is another sense of the term that they are; according to this second sense of the term, to claim that cognitive abilities are dispositions means that abilities are *character traits*, or habitual behaviors that the agent tends to exhibit. A strong indication that this is how we should understand the dispositional nature of cognitive abilities has to do with the fact that abilities can only be acquired and sustained through practice, whereas dispositions, in the other meaning of the term, can be possessed by an entity even if they are never actually manifested (e.g., a vase may be fragile even if it has never been broken). As we shall see below, Greco appears to concur with this understanding of abilities as he claims they are the stable traits of the agent’s cognitive *character*; a behavior can be *in character* only if it is habitually manifested. See also (Greco 2010, 150). I am thankful to the referee for pressing this point.

become epistemically responsible in employing a process is that it be a disposition or habit of hers.<sup>10,11</sup>

So to summarize what we have gathered from the three examples above, in cases of knowledge, we want one's beliefs to be responsive to the facts. Accordingly, we claimed that only *prima facie* cognitive processes can be knowledge-conducive, but not all of them will do. The process must be a cognitive ability, meaning that the agent will be able to be conscientious and thereby subjectively justified in employing it. And in the absence of positive reasons (that are accessible through reflection alone) for the reliability of the cognitive process, this last condition can be satisfied in a realistic way only if the relevant process is a normal disposition or habit of the agent such that it can become part of (or be integrated into) the rest of her cognitive system.

Notably then, the general idea, which all the above considerations are alluding to, is that for a reliable process to be knowledge-conducive it must be a cognitive ability. This idea has also appeared in the literature as the *ability intuition on knowledge* and can be summarized by stating that *knowledge is belief that is true in virtue of cognitive ability*.<sup>12</sup> Now the end result of the above considerations is that they fill in the details of which reliable processes may plausibly count as cognitive abilities by demonstrating that it is only normal, dispositional or habitual cognitive processes that the agent can be subjectively justified in employing (and thereby able to gain knowledge from).

Now, building on considerations very similar to the above ones, Greco (1999) has proposed a virtue reliabilist account of knowledge, which emphasizes that when we assess whether some agent knows, we shouldn't be focusing on the reliability of isolated (cognitive) belief-forming processes, but on the reliability of the overall agent, conceived of as a stable, interconnected system of such belief-forming processes.<sup>13</sup> It is this interwoven totality of

---

<sup>10</sup> I here say 'the only realistic way' because we can imagine, for instance, a case of a benevolent mentalist who hypnotizes the agent to trust a newly acquired process, and trust it only in the appropriate conditions (thereby allowing him to be epistemically responsible in employing it), despite the fact that the process is not a disposition of hers.

<sup>11</sup> For further discussion of the above intuitions on the Serendipitous Brain Lesion and Careless Math Student cases see (Greco 1999) and (Greco 2010, 149-155). For the discussion of similar thought experiments and intuitions see (Bonjour 1980), (Goldman 1986), (Lehrer 1990), and (Plantinga 1993b).

<sup>12</sup> The idea that knowledge must be grounded in cognitive abilities can be traced back to the writings of Sosa (1988; 1993) and Plantinga (1993a). For more recent approaches to this intuition, see Greco (1999; 2004; 2007) and Pritchard (2009; *forthcoming*; 2010a; 2010b).

<sup>13</sup> Any theory of knowledge that places in its center the ability intuition on knowledge will fall under the general trend of *Virtue Reliabilism* (abilities are normally understood as virtues

cognitive abilities that give rise to one's sense of epistemic self (or in Greco's terms to one's 'cognitive character'), and which should be the focus of our epistemic assessment.

To make apparent the motivation for this change of epistemic focus, remember that in order to avoid the Humean problematic we must accommodate subjective justification in a way that does not involve knowledge or even beliefs about reliability. At this point, Greco (1999, 289) suggests that a promising strategy for doing so is to claim that "a belief *p* is subjectively justified for a person *S* (in the sense relevant for having knowledge) if and only if *S*'s believing *p* is grounded in the cognitive dispositions that *S* manifests when *S* is thinking conscientiously" (i.e., when *S* is motivated to believe what is true). In this way, the agent will employ his reliable cognitive processes only in circumstances that have not been problematic in the past, *and* he will be able to do so without even having *any* beliefs about their reliability.<sup>14</sup> Greco, then, goes on to further claim that the dispositions/habits that a person manifests when she is thinking conscientiously intertwine with each other and give rise to what we may call one's *cognitive character* (Greco 1999, 290). So, overall, "a belief *p* has a positive epistemic status for a person *S* just in case *S*'s believing *p* results from the stable and reliable dispositions that make up *S*'s cognitive character" (*ibid.*, 287-8).

---

and *vice versa*). To accentuate the features of his account, Greco calls his view *Agent Reliabilism*, but it is clearly a version of *Virtue Reliabilism*—one that emphasizes the importance of the overall agent in the manifestation of the relevant intellectual virtues. For alternative, robust as well as weaker formulations of *Virtue Reliabilism* see Sosa (1993; 2007) and Pritchard (*forthcoming*; 2010a; 2010b), respectively.

<sup>14</sup> The fact that people manifest highly specific, finely tuned dispositions to form their beliefs in certain ways but not in others amounts to an implicit awareness of the reliability of those dispositions.

For example suppose that it seems visually to a person that a cat is sleeping on the couch, and on this basis she believes that there is a sleeping cat on the couch. Suppose also that this belief manifests a disposition that the person has, to trust this sort of experience under these sorts of conditions, when motivated to believe the truth. Now, suppose that much less clearly, it seems visually to the person that a mouse has run across the floor. Not being disposed to trust this kind of fleeting experience, the person refrains from believing until further evidence comes in. The fact that the person, properly motivated, is disposed to trust one kind of experience but not the other, constitutes sensitivity on her part that the former is reliable. There is a clear sense in which she takes the former experience to be adequate to her goal of believing the truth, and takes the latter experience not to be. And this is so even if she has no beliefs about her goals, her reliability, or her experience (Greco 1999, 290 ) .

A similar argument can be found in (Sosa 1993, 60-63).

In this way, Greco can do away both with strange and fleeting processes. Strange processes cannot be part of the agent's cognitive character because they are not the kind of processes that a conscientious agent would employ. Fleeting processes are also excluded. First, because they are not dispositions or habits—so they cannot really count as character traits. And, second (I may add), because, in the absence of reasons to believe that the relevant process is reliable, it is only dispositions or habits that one can become aware they are unreliable in certain circumstances and, so—without relying on any beliefs about their reliability—use them conscientiously in the rest of the circumstances.<sup>15</sup>

So in order to gain knowledge on the basis of a process the agent must be able to employ a conscientious attitude towards that process. And in order for that to be the case, the relevant process must be a cognitive ability of the agent, meaning that it must have been *integrated* into the agent's cognitive character. Now, despite the previous points on the importance of the normality and dispositionality of the relevant process in order for it to count as a genuine part of one's cognitive character, Greco attempts to further accentuate and shed some light on the *integrated* nature of our cognitive characters by noting that the process of "cognitive integration is a function of cooperation and interaction, or cooperative interaction with other aspects of the cognitive system" (2010, 152). So, how exactly should we think about the required conditions for a process to count as knowledge-conducive?

In general, every knowledge-conducive process must be a cognitive ability such that the agent will be subjectively justified in employing it, which requires that the process be integrated into the agent's cognitive character by cooperatively interacting with it. Accordingly, we may say that the only necessary and sufficient condition for a process to count as knowledge-conducive is that it cooperatively interacts with the rest of the agent's cognitive character. Now, apparently, this makes the normality and dispositionality criteria seem redundant—which strictly speaking they are—but they may still have a role to play; normality and dispositionality of the relevant process seem to be *practical preconditions* for the agent to be able to

---

<sup>15</sup> Apart from the example given in the previous footnote, Greco has not attempted to provide an account of how the process of subjective justification works. I assume, however, that he wouldn't reject this falsificationist approach, as I cannot see how else subjective justification could be accommodated in an externalist way.

cooperatively interact with it. The extent, however, to which each one of these criteria may need to be satisfied will differ from case to case. An agent, for example, may be subjectively justified in employing a process in the appropriate conditions not because it is a normal disposition of hers but because a benevolent mentalist hypnotized her to do so. In most realistic cases, however, normality and dispositionality will still have a significant guiding effect. The decisive effect of cognitive integration, however—on the basis of which the agent can be conscientious and thereby subjectively justified—will only be ensured if the agent's cognitive character mutually interacts with the relevant process. So, all we need to accept that a process is knowledge-conducive is that it be integrated into the agent's cognitive character by cooperatively interacting with it. On the basis of this mutual interaction with the rest of his cognitive system the agent will be able to employ the relevant process conscientiously by merely lacking any beliefs that it is unreliable (at least not in the circumstances in which he employs it), or if the employment is involuntary, conscientiously accept its deliverances when he lacks any beliefs that the conditions that gave rise to them were unreliable.

Moreover, as we saw previously in the discussion of the three examples, this process of cognitive integration gives rise to a coherentist effect both on the level of processes (how the beliefs are generated) and on the level of content (how the beliefs themselves combine). Also, it ensures that at least directly related belief-generating mechanisms and their resulting beliefs will be consistent.

Overall then, the epistemic importance of cognitive integration is that it allows epistemic agents to satisfy the condition of subjective justification in a minimalist way, which is nevertheless sufficient for acquiring knowledge even if one cannot—not even in principle—offer any explicit reasons in favor of one's beliefs. Stated explicitly, this minimalist condition of subjective justification is that conscientious epistemic agents ought to accept the deliverances of, and employ their cognitive abilities, only when they lack any doubts that they are unreliable, given the conditions they employ them in.<sup>16</sup> In

---

<sup>16</sup> As an anonymous referee has pointed out I should make clear that this condition should be restricted to reliable processes. We should not allow, for example, to an agent who forms his beliefs on the basis of astronomical considerations or wishful thinking to count as subjectively justified merely by lacking any doubts about the unreliability of his belief-forming processes.



other words, cognitive integration allows epistemic agents to be conscientious in the sense that in cases where there is something wrong with the way they form their beliefs they will be able to spot this and respond appropriately. Otherwise, they can go on formulating their beliefs, without worrying whether they can actually offer reasons for every single one of them or for their reasons for holding them.<sup>17</sup>

### 2.3. *Cognitive integration in epistemology and philosophy of mind*

In the introduction, I mentioned that the topic of cognitive integration has been explored in epistemology also in reference to the possibility of cognitive extension. The general idea is that the knowledge-conducive cognitive abilities and their relation to one's cognitive character as discussed within virtue epistemology is particularly apt for an interpretation along the lines suggested by the hypothesis of extended cognition. To make this clear it should be helpful to repeat, for the last time, what are the important features of a knowledge-conducive belief-forming process.

In general, the process must be a cognitive ability. In order for that to be the case the process must be a cognitive process. This will guarantee the correct direction of fit between the belief and the fact. Also, we want the belief-forming process to be objectively reliable, where a reliable process is

---

Given, however, that the epistemic agent must be conscientious (i.e., motivated to believe what is true) this qualification is actually redundant. If the agent is motivated to believe what is true, he will not employ astronomical considerations or wishful thinking because he will have noticed that such processes were notably unreliable in the past. For the same reason, they won't even be parts of his (conscientious) cognitive *character*.

<sup>17</sup> Further to footnotes 14 and 15, in providing this sort of account of subjective justification, I have relied for the most part on phenomenological intuitions about how we seem to go about our beliefs in everyday life. Nevertheless, such phenomenological intuitions seem to already entertain a certain degree of scientific support. Specifically, within cognitive psychology, there have been several studies indicating that subjects engage in analytic reasoning only when they experience the metacognitive effect of the lack of 'fluency':

"Fluency is not a cognitive operation in and of itself but, rather, a feeling of ease associated with a cognitive operation, it can be generated by nearly any form of thinking. If a percept is blurry, we are aware that it was hard to see. If a word is phonemically irregular, we recognize the challenge in processing it. We know whether we had to struggle to bring a memory to mind and whether we had a hard or easy time solving a riddle. Because the metacognitive experience of fluency can be generated by so many cognitive processes and is nearly effortless to access, it can serve as a cue toward judgments in virtually any situation". (Oppenheimer *forthcoming*).

For an overview on the metacognitive feeling of fluency see (Oppenheimer *forthcoming*), (Alter & Oppenheimer 2007) and (Unkelbach & Greifeneder 2013).

one that tends to produce true rather than false beliefs. Recall, however, that according to reliabilism, the agent does not need his evidence to be necessarily reliable, such that he can be internally justified in holding his belief; if forming a belief on a certain kind of evidence constitutes a reliable belief-forming process, it does not matter that one's evidence is only contingently reliable; the agent, on his part, does not need to know or even have any beliefs about the reliability of his way of forming beliefs. Instead, the agent can be subjectively justified simply by forming his belief on the basis of a process that is integrated into his cognitive character, which he employs when he is thinking conscientiously. Now, in order for a process to be a candidate for inclusion to the agent's conscientious cognitive character, we noted that it will probably have to be neither strange nor fleeting. In most realistic scenarios the process (1) will have to be normal so that the agent won't reject it when conscientious and (2) will have to be a disposition or a habit of the agent, because (barring scenarios such as the mentalist case) it is only dispositions or habits that one can become aware they are unreliable in certain circumstances, and so—without relying on any beliefs about their reliability—be able to employ them conscientiously in the rest of the circumstances. As we further noted, however, even though normality and dispositionality will, in most cases, be practical preconditions, they are neither necessary nor sufficient for a process to count as integrated into the agent's cognitive character. Instead, the only thing that is required is that the process be integrated into the agent's cognitive character, by engaging in cooperative interaction with the rest of the agent's cognitive system. Accordingly, no matter what the practical preconditions for this interactive process to be achieved are, once it is in place, it will guarantee both that the relevant belief-forming process is a cognitive process, and that it is indeed part of the agent's cognitive character such that he can be conscientious in employing it. So putting all the above points together: a belief-forming process counts as a cognitive ability and thereby as knowledge-conducive if and only if it is a reliable belief-forming process that is integrated into the agent's cognitive character, on the basis of a process of cooperative interaction with it.<sup>18</sup>

---

<sup>18</sup> Many externalist epistemologists would reject the above biconditional on the grounds that in order for a process to be knowledge-conducive it should also be safe (where a safe process is one that could not have easily been wrong). Consider for example *Anti-Luck Virtue Epistemology*: *S* knows that *p* if and only if *S*'s safe belief that *p* is the product of her relevant cognitive

Now to see why we might get the impression of a close fit between virtue epistemology and the hypothesis of extended cognition, here are three common-sense functionalist criteria, which Clark (2010) suggests must be satisfied by non-biological candidates in order to be included into an individual's cognitive system:

- 1) "That the process be reliably available and typically invoked".

That is, the agent should habitually and easily invoke the external resource. In other words, its employment must be a disposition/habit of the agent's overall cognitive mechanism.

- 2) "That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny. [...] It should be deemed about as trustworthy as something retrieved clearly from biological memory".

That is, the information in the resource must be *regarded* as normal and reliable and not be necessarily reliable. It suffices that its employment result into an equally trustworthy belief-forming process as the one of forming beliefs on the basis of one's own biological memory.<sup>19</sup>

At this point, however, one might object that being reliable is not the same as being *trustworthy* (i.e., being *regarded* as reliable). But, in response, notice first that Clark identifies the notion of trustworthiness of a process with the idea of being "more-or-less automatically endorsed" or in other words "not usually subject to critical scrutiny". That is, the target process must not have been (for the most part) problematic in the past. Moreover, the processes under consideration are also supposed to be cognitive dispositions or habits of the agent that he has repeatedly employed in the past, and so had they been problematic the agent would have noticed that and responded appropriately. Accordingly, a *trustworthy* belief-forming process in Clark's account, will be one that tends to produce true rather than false beliefs, which

---

*abilities (such that her safe cognitive success is to a significant degree creditable to her cognitive agency) (Pritchard forthcoming, 20). Again, in (Pritchard 2010a, 76) we can read: "knowledge is safe belief that arises out of the reliable cognitive traits that make up one's cognitive character, such that one's cognitive success is to a significant degree creditable to one's cognitive character". For a defense of the claim that the safety condition is not necessary for virtue reliabilism to account for knowledge see (Palermos forthcoming)*

<sup>19</sup> That is, the process does not need to be, due to underlying logical or quasi-logical relations, 100% reliable. Notice that memory is supposed to be reliable even though one may misremember.

is to say that it will be objectively reliable in the virtue reliabilist's sense. What the agent will deem reliable will be that which is objectively reliable, i.e., that which has *not* been (for the most part) problematic in the past.

Furthermore, notice this negative way of deeming processes reliable with which Clark concurs (i.e., that a trustworthy process is one that is *not usually* subject to critical scrutiny such that it is more-or-less automatically endorsed). What this means is that the agent does not need to have any beliefs about why or whether his belief-forming process is trustworthy; it suffices that it has not repeatedly caught his negative attention in the past. This is in good agreement with the proposed minimalist understanding of subjective justification according to which one does not need to rely on any beliefs but simply on one's motivation to believe the truth. For example, one will trust one's vision in appropriate circumstances, just because vision has not been notably problematic in the past (in those circumstances). By being motivated to believe the truth one will thereby employ the belief-forming process that has not in the past (notably) failed to be conducive towards that end, and crucially, one will do so without even thinking about it.

- 3) "That information contained in the resource should be easily accessible as and when required".

That is, the agent must be able to employ it as if it was part of his organismic cognitive mechanism. In other words, the resource must be *integrated* into the agent's overall cognitive mechanism.

So we see that the same features of a process that epistemologists deem important in order for a process to be knowledge-conducive are required by a common-sense functionalist understanding of cognition in order for a process to count as part of one's mind. This is a promising observation.

Notice, however, that even if there were no such close fit between these broad features, we would still be able to show that the two theories are essentially connected. The reason, as we noted before, is that some of the above features (e.g., normality and dispositionality of the process) may be conducive towards the process being knowledge-conducive, but they are neither necessary nor sufficient for that end. Instead, the only requirement is that the process be integrated into the agent's cognitive character.

Now in relation to this, philosophy of cognitive science has recently started shifting its focus away from the above common-sense functionalist criteria for including an external resource into one's cognitive system. Instead, Chemero (2009), Froese et al. (2013) and I (Palermos 2014) have suggested that the only requirement for an external element to count as a constitutive part of the agent's cognitive system is that it be non-linearly related to the rest of the agent's cognitive system. The motivation for this is that, according to dynamical systems theory, these non-linear relations give rise to an *overall non-decomposable system* that consists of all the contributing parts. And two reasons for postulating the overall system are that these non-linear interactions (1) give rise to new systemic properties that belong only to the overall system and to none of the contributing systems alone (therefore we have to postulate the overall extended system) and (2) prevent us from decomposing the two systems in terms of distinct inputs and outputs from the one subsystem to the other (therefore we cannot but postulate the overall system).<sup>20</sup> What is even more interesting to our present purposes, however, is that just as Greco holds that cognitive integration is a matter of interaction and cooperation between cognitive processes, so those non-linear relations that allow us to talk about integration within philosophy of mind and cognitive science emerge only on the basis of *cooperative feedback loops* between the contributing elements of the overall system.

Therefore both in epistemology and philosophy of mind and cognitive science the same criterion (cooperative interaction with the rest of the agent's cognitive system) is required for a process to be integrated into an agent's cognitive system and thereby count as knowledge-conducive. This, however, should not really come as a surprise. Given that virtue reliabilism holds that knowledge must be the product of cognitive ability (however that ability may be realized) and that the hypothesis of extended cognition sets out to reveal which processes can count as cognitive abilities (wherever they may be located), this close fit between the two theories seems to be as it should be.

The conclusion that follows, then, is that there is no principled theoretical bar disallowing extended belief-forming processes from counting as knowledge-conducive cognitive abilities. Given that virtue reliabilism

---

<sup>20</sup> For a detailed explanation of why the existence of non-linear relations that arise out of the mutual interactions between agents and their artifacts ensures the existence of extended cognitive systems see (Palermos 2014).

makes no specifications as to whether knowledge-conducive cognitive abilities should be located within the agent's head, then, provided that the condition of cognitive integration is met, the epistemic agent may extend his knowledge-conducive cognitive character beyond his organismic cognitive abilities by incorporating epistemic artifacts to it.

### 3. IS THIS NECESSARY FOR EPISTEMOLOGY AND WHAT ARE THE RAMIFICATIONS?

Obviously, the possibility of knowledge-conducive cognitive characters that may nevertheless be extended beyond our organismic cognitive capacities can generate interesting ramifications both for traditional and social epistemology, which may allow us to think about knowledge in new ways. Focusing, however, on the integrated nature of our extendable cognitive characters may not only be important for moving forward, but also necessary for accounting for knowledge as we already think about it.

According to Greco (1999, 287), in addition to one's organismic cognitive abilities of the brain/central nervous system, a person's cognitive character may also consist of "acquired skills of perception and acquired methods of inquiry including those involving highly specialized training or even advanced technology". The reason for this move is that we need to account for advanced cases of knowledge where one's believing the truth is the product of the operation of epistemic artifacts such as telescopes, microscopes, tactile visual substitution systems and so on.<sup>21</sup> The problem, however, is that in the traditional conception, cognition takes place strictly within the agent's head and so artifacts cannot be parts of one's *cognitive* character.

One way to sidestep this problem for virtue reliabilism, could be to claim that, in such cases, it is merely the agent's training and skill of using the artifact, as mirrored in the agent's neural/bodily architecture, that is the most salient factor in the causal explanation of the agent's cognitive success (i.e., believing the truth). Notice, however, that when an agent employs an epistemic tool, his true belief arises as the product of the *interaction* between his internal processes and the artifact. According to dynamical systems

---

<sup>21</sup> See Bach-y-Rita and Kercel (2003) for a recent review on tactile visual substitution systems.

theory, then, the cognitive process that allows the agent to detect the truth is not merely ‘aided’ or ‘assisted’ by the artifact but is, instead, constituted by it as it arises out of the ongoing mutual interaction between the agent and the artifact.<sup>22</sup> Therefore, in a causal explanation of how the agent acquired his true belief, it will be impossible to disentangle the agent’s training and skill of using the artifact from his actual engagement with it.<sup>23</sup>

But even if such decomposition were possible, notice in addition that the part of the process that allows the agent to detect the truth, or in other words to be sensitive to the facts, is the external component. To illustrate this, consider, on one hand, an untrained agent in possession of a properly working artifact. In that case, it is obvious that even though the agent will initially be unable to form any (true or false) beliefs, eventually—provided that he gains sufficient experience such that he can interact with it—not only will he form beliefs, but he will also reliably enjoy cognitive success. On the other hand, think about a well-trained agent, but in possession of a faulty artifact. In this case, despite the agent’s excellent internal skills, it is evident that he would be unable to reach any (non-lucky) true beliefs, no matter how much he tried. It therefore seems that in such cases the most (and maybe the only) significant factor that explains the truth-status of the agent’s belief is the epistemic artifact. In other words, since the agent’s belief is true in virtue of the artifact, the virtue reliabilist must account for it being part of the agent’s cognitive system. Given, however, that cognition is normally supposed to take place within the agent’s head, virtue reliabilists can only account for such cases by wedding their view to the hypothesis of extended cognition. Accordingly, combining the extended cognition hypothesis with virtue reliabilism on the basis of their close fit does not seem to be just an available option for epistemologists, but also necessary for dealing with advanced cases

---

<sup>22</sup> It should be here noted that not every case of the employment of an artifact is a case of cognitive extension, but only when the agent mutually interacts with it. For an objective criterion of constitution and on what may count as a genuine case of cognitive extension, see (Palermos 2014).

<sup>23</sup> Remember that according to virtue reliabilism and the underlying ability intuition on knowledge, knowledge is belief that is true in virtue of cognitive ability, where, according to Greco, “in virtue of” must be understood in causal explanatory terms. Even though several proponents of virtue reliabilism agree on this general causal-explanatory understanding of the view, there is disagreement on whether the relevant cognitive ability should be the “most salient” (Greco 2010) or merely a “significant” (Pritchard 2010b) factor in the causal explanation of how the agent acquired his true belief.

of knowledge where the latter is the product of the employment of epistemic artifacts that the agent mutually interacts with.

Apart from the necessity of introducing the extended cognition hypothesis within epistemology, however, this move can also generate interesting ramifications, especially for social epistemology. For instance, it can lead to the further claim that there could be cognitive characters that do not just extend beyond an agent's organismic capacities, but which are instead *distributed* amongst *several agents* along with their epistemic artifacts. The hypothesis of distributed cognition, which has been developed in parallel to the hypothesis of extended cognition (Hutchins 1995, Theiner et al. 2010, Sutton et al. 2008, Wilson 2005, Heylighen et al. 2007), differs to the latter position only in that this time the cognitive system extends to include epistemic artifacts as well as other agents. And interestingly, most proponents of the view (Sutton et al. 2008, Theiner et al. 2010, Wegner 1985, Tollefsen & Dale 2011) point out again that it is the existence of non-linear cooperative interactions between the contributing members and their artifacts that is the criterion by which we can judge whether we have an *integrated distributed cognitive system*. Accordingly, there could be knowledge-conducive cognitive characters, which may nevertheless be distributed.

This is an interesting possibility, because it can allow us to combine an individualistic approach to knowledge, such as virtue reliabilism, with the hypothesis of distributed cognition in order to account for *epistemic group agents*: Groups of individuals who exist and gain knowledge in virtue of a shared, common cognitive character that mainly consists of a distributed cognitive ability—a collective cognitive ability that emerges out of the members' mutual (socio-epistemic) interactions and which is not reducible to the cognitive abilities possessed by the individual members, thereby allowing us to speak of a group agent in itself. This is important, because by recognizing a group of people as a self-standing agent in itself, we can then use an individualistic approach to knowledge to account for knowledge that is collectively produced and which is, thereby, distinctively social. In other words we can make sense of the claim that *p* is known by *S* (the group agent), even though it is not known by any individual alone.<sup>24</sup>

---

<sup>24</sup> For a more detailed explanation of how virtue reliabilism may be applied to epistemic group agents see (Palermos & Pritchard 2013).



Remarkably, such group agents have already started being studied within cognitive science. Consider for example transactive memory systems (TMSs)—i.e., groups of two or more individuals who collaboratively encode, store and retrieve information. The reason why TMSs are good candidates for distributed cognitive systems—and thereby for epistemic group agents—is that, as Sutton et al. observe (2008), such systems are likely to involve skillful interactive simultaneous coordination of people who can thereby count as a single integrated cognitive system. Therefore, we can use TMSs in order to “conceptualize how people in close relationships may depend on each other for acquiring, remembering, and generating knowledge” (Wegner et al. 1985, 253):

Ordinarily psychologists think of memory as an individual’s store of knowledge, along with the processes whereby that knowledge is constructed, organized, and accessed. [...] With transactive memory we are concerned with how knowledge enters the dyad, is organized within it, and is made available for subsequent use by it (*ibid.*, 256).

Apart from incorporating already existing research from cognitive science, however, the combination of virtue reliabilism with the hypotheses of extended and distributed cognition can generate new avenues for research, some of which have for a long time been inaccessible. An interesting example is the intersection between epistemology and the field of history and philosophy of science. These two intimately related fields have so far been at odds—an awkward situation owing to the fact that the former discipline has traditionally been individualistic whereas the latter has for the most part been socially oriented (hardly anyone could deny the social nature of the scientific process, especially after the publication of Kuhn’s *The Structure of the Scientific Revolutions*, in 1962). The present approach, however, could now provide a useful link between the two fields. Science is primarily performed by individual scientists employing their hardware and software epistemic artifacts or by *research teams* operating within scientific labs that are uniquely tailored to fit their purposes. Accordingly, the concepts of extended cognitive characters and epistemic group agents could become very handy for a mainstream epistemological analysis of the scientific progress. As Giere and Moffat (2003, 308) note in their discussion of the scientific revolution of the 16th century,

“No ‘new man’ suddenly emerged sometime in the sixteenth century....The idea that a more rational mind...emerged from darkness and chaos is too complicated a hypothesis” [Latour 1986, 1]. We agree completely. Appeals to cognitive architecture and capacities now studied in cognitive sciences are meant to explain how humans with normal human cognitive capacities manage to do modern science. One way, we suggest, is by constructing distributed cognitive systems that can be operated by humans possessing only the limited cognitive capacities they in fact possess.

#### 4. CONCLUSION

The topic of cognitive integration is a far-reaching multidisciplinary theme that touches upon a wide range of disparate questions with the potential of bringing several research areas together. Within epistemology, the phenomenon of cognitive integration can reveal how knowers may be subjectively justified despite the absence of positive reasons in favor of their true beliefs, pointing towards a minimalist yet informative epistemic norm: provided that one’s belief-forming process has been integrated into one’s cognitive system/ character and that one is motivated to believe what is true, one can be justified in holding the resulting belief merely by lacking any doubts that his way of forming his belief, or that the belief itself, is inappropriate. Within philosophy of mind, cognitive integration can explain how it is possible to extend our cognitive capacities beyond our organisms to the artifacts we employ or other individuals we may interact with. And combining philosophy of mind with epistemology by using the process of cognitive integration as the connecting point provides the necessary means to account for advanced cases of knowledge where the known belief is true in virtue of the operation of epistemic artifacts or even the activity of collaborative groups. Quite likely, however, such an account of knowledge won’t be valuable just within epistemology, philosophy of mind and their intersection. Focusing on the essentially technological and collaborative nature of the scientific process, such an approach to knowledge could finally provide a strong link between the related but so far persistently isolated fields of philosophy of mind, philosophy of science, and epistemology.

## Acknowledgments

I am very thankful to John Greco and an anonymous referee for detailed comments on previous versions of this paper. Research in the area of this paper was carried out as part of the AHRC-funded 'Extended Knowledge' research project (AH/J011908/1).

## References

- Alter, L., A. & Oppenheimer, D., M. (2007). 'Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning'. *Journal of Experimental Psychology*. Vol. 136, No. 4, 569-576.
- BonJour, L. (1980). "Externalist Theories of Empirical Knowledge," *Midwest Studies in Philosophy*, 5: 53–73.
- (1985). *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.
- (2002). 'Internalism and Externalism', *Oxford Handbook of Epistemology*, (ed.) P. Moser, 234-64, Oxford: Oxford University Press.
- Brewer, F. W. & Lambert, B. L. (2001). 'The Theory Ladeness of Observation and the Theory-Ladeness of the Rest of the Scientific Process'. *Philosophy of Science*, Vol. 68, No. 3, Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I: Contributed Papers, pp. S176-S186.
- Clark, A., & Chalmers, D. (1998). 'The Extended Mind'. *Analysis* 58, no. 1: 7-19.
- (2008). *Supersizing The Mind*. Oxford University Press.
- (2010). 'Memento's Revenge: The Extended Mind, Extended'. In the *Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT press.
- Chisholm, R. M. (1977). *Theory of Knowledge* (2<sup>nd</sup> ed.), Englewood Cliffs, NJ: Prentice-Hall.
- Churchland, P. M. (1979). *Scientific Realism and the Plasticity of the Mind*. Cambridge: Cambridge University Press.
- (1988). 'Perceptual Plasticity and Theoretical Neutrality: A Reply to

- Jerry Fodor', *Philosophy of Science* 55: 167-187.
- (1989). *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*. Cambridge: The MIT Press.
- Estany, A. (2001). 'The Thesis of Theory-Laden Observation in the Light of Cognitive Psychology'. *Philosophy of Science*, Vol. 68, No.2 (Jun ., 2001), pp. 203-217.
- Feyerabend, P. K. (1975). *Against Method: Outline of an anarchistic theory of knowledge*.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- (1984). 'Observation Reconsidered', *Philosophy of Science*, 51: 23-43.
- (1988). 'A Reply to Churchland's 'Perceptual Plasticity and Theoretical Neutrality'', *Philosophy of Science* 55: 188-198.
- Froese, T., Gershenson, C., and Rosenblueth, D., A. (2013). 'The Dynamically Extended Mind'. <http://arxiv.org/abs/1305.1958>
- Fuller, S. (2012). 'Social Epistemology: A Quarter-Century Itinerary', *Social Epistemology: A Journal of Knowledge, Culture and Policy* 26, 267-83.
- Giere, R. & Moffat, B. (2003). 'Distributed Cognition: Where the Cognitive and the Social Merge'. *Social Studies of Science*. 33/2, pp. 1-10.
- Goldman, A. (1986). *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.
- Greco, J. (1999). 'Agent Reliabilism', in *Philosophical Perspectives* 13: Epistemology (1999). James Tomberlin (ed.), Atascadero, CA: Ridgeview Press, pp. 273-296.
- (2004). 'Knowledge As Credit For True Belief', in *Intellectual Virtue: Perspectives from Ethics and Epistemology*. M. DePaul & L. Zagzebski (eds.), Oxford: Oxford University Press.
- (2007) 'The Nature of Ability and the Purpose of Knowledge', *Philosophical Issues* 17, pp. 57- 69.
- (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.
- Hanson, N. R. (1961). *Patterns of Discovery*. Cambridge: Cambridge University Press.
- (1969). *Perception and Discovery; An Introduction to Scientific Inquiry*. San Francisco: Freeman, Cooper.

- Heylighen, F., Heath, M., Van Overwalle, F. (2007). 'The Emergence of Distributed Cognition: A Conceptual Framework'. In *Proceedings of collective intentionality IV (2004)*, Volume: IV, Publisher: University of Siena.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
- Oppenheimer, D., M. (2008). 'The Secret Life of Fluency'. *Trends in Cognitive Science*. 12(6):237-41.
- Palermos, O. (2011). 'Belief-Forming Processes, Extended', *Review of Philosophy and Psychology* 2, 741-65.
- (2014). 'Loops, Constitution, and Cognitive Extension', *Cognitive Systems Research*, 27: 25-41.
- (forthcoming). 'Could Reliability Naturally Imply Safety?', *European Journal of Philosophy*, DOI: 10.1111/ejop.12046.
- Palermos, S. O., & Pritchard, D. (2013). 'Extended Knowledge and Social Epistemology', *Social Epistemology Review and Reply Collective* 2 (8): 105-120.
- Plantinga, A. (1993a). *Warrant and Proper Function*. New York: Oxford University Press.
- (1993b). *Warrant: The Current Debate*, Oxford: Oxford University Press.
- Pappas, G. (2005). 'Internalist versus Externalist Conceptions of Epistemic Justification', *Stanford Encyclopaedia of Philosophy*, (ed.) E. Zalta, <http://plato.stanford.edu/entries/justep-intext/>
- Poston, T. (2008). 'Internalism and Externalism in Epistemology', *Internet Encyclopaedia of Philosophy*, (eds.) B. Dowden & J. Fieser, [www.iep.utm.edu/int-ext/](http://www.iep.utm.edu/int-ext/)
- Pritchard, D. (forthcoming). 'Anti-Luck Virtue Epistemology', *Journal of Philosophy*.
- (2009). *Knowledge*, London: Palgrave Macmillan.
- (2010a). 'Knowledge and Understanding', in A. Haddock, A. Millar & D. H. Pritchard, *The Nature and Value of Knowledge: Three Investigations*, Oxford: Oxford University Press.
- (2010b). 'Cognitive Ability and the Extended Cognition Thesis'. *Synthese*.

- Pryor, J. (2001). 'Highlights of Recent Epistemology', *British Journal for the Philosophy of Science* 52, 95-124.
- Sosa, E. (1988). 'Beyond Skepticism, to the Best of our Knowledge'. *Mind*, New Series, vol. 97, No.386, pp. 153-188
- (1993). 'Proper Functionalism and Virtue Epistemology'. *Nous*, Vol. 27, No. 1, 51-65.
- (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, Oxford: Clarendon Press.
- Steup, M. (1999). 'A Defense of Internalism', *The Theory of Knowledge: Classical and Contemporary Readings* (3<sup>rd</sup> Ed.), (ed.) L. Pojman, 310-21, Belmont, CA: Wadsworth.
- Sutton, J., Barnier, A., Harris, C., Wilson, R. (2008). 'A conceptual and empirical framework for the social distribution of cognition: The case of memory'. *Cognitive Systems Research*, Issues 1-2, pp. 33–51.
- Theiner, G. & Allen, C. & Goldstone, R. (2010). 'Recognizing Group Cognition'. *Cognitive Systems Research*, Vol. 11, Issue 4, pp. 378-395.
- Tollefsen, D. & Dale, R. (2011). "Naturalizing joint action: A Process-based approach". *Philosophical Psychology* 25 (3):385-407.
- Unkelbach, C. & Greifeneder, R. (2013). 'A general model of fluency effects in judgment and decision making'. In C. Unkelbach & R. Greifeneder (Eds.), *The Experience of Thinking* (pp. 11-32). New York: Psychology Press.
- Wegner, M., Giuliano, T., Hertel, P. (1985). 'Cognitive interdependence in close relationships'. In W. J. Ickes (Ed.), *Compatible and incompatible relationships* (pp. 253–276). New York: Springer-Verlag.
- Wilson, R., A. (2005). 'Collective Memory, Group Minds, and the Extended Mind Thesis'. *Cognitive Processing*, Vol. 6, Issue 4, pp. 227-236.